Achieving Vanishing Rate Loss in Decentralized Network MIMO

Antonio Bazco-Nogueras[*†], Lorenzo Miretti[†], Paul de Kerret[†], David Gesbert[†], Nicolas Gresset[*]

[*]Mitsubishi Electric R&D Centre Europe (MERCE), 35700 Rennes, France, {a.bazconogueras,n.gresset}@fr.merce.mee.com
[†]Communication Systems Department, EURECOM, 06410 Biot, France, {bazco,miretti,dekerret,gesbert}@eurecom.fr

Abstract—In this paper[1], we analyze a Network MIMO channel with 2 Transmitters (TXs) jointly serving 2 users, where each TX has a different multi-user Channel State Information (CSI), potentially with a different accuracy. Recently it was shown the surprising result that this decentralized setting can attain the same Degrees-of-Freedom (DoF) as its genie-aided centralized counterpart in which both TXs share the best-quality CSI. However, the DoF derivation alone does not characterize the actual rate and the question was left open as to how big the rate gap between the centralized and the decentralized settings was going to be. In this paper, we considerably strengthen the previous intriguing DoF result by showing that it is possible to achieve asymptotically the same sum rate as that attained by Zero-Forcing (ZF) precoding in a centralized setting endowed with the best-quality CSI. This result involves a novel precoding scheme which is tailored to the decentralized case. The key intuition behind this scheme lies in the striking of an asymptotically optimal compromise between i) realizing high enough precision ZF precoding while ii) maintaining consistent-enough precoding decisions across the non-communicating cooperating TXs.

I. INTRODUCTION

Joint transmission in wireless networks is known to bring multiplicative improvements in network rates only under the assumption of perfect CSI [1]. The study of how imperfect or quantized CSI at the TXs (CSIT) affects the performance has focused on the assumption that the imperfect information is *perfectly shared* between the non-colocated transmitting antennas [1], [2]. However, this assumption may not be adapted to many applications within the upcoming wireless networks use cases, such as Ultra-Reliable Low-Latency Communication (URLLC) or heterogeneous backhaul deployments. As a result, there is a clear interest in looking at the scenario in which each TX may have a different information about the channel, denoted as *Distributed CSIT* setting [3].

We focus in this paper on a particular sub-case of the Distributed CSIT setting, so-called Distributed Network MIMO, wherein the TXs have access to all the information symbols of the users (RXs), yet do not share the same CSIT [4]. This model arises in presence of caching [5] and Cloud-RAN with high mobility [6], in which latency constraints impede efficient CSIT sharing within the channel coherence time. The DoF of this scenario has been studied in previous works. Specifically, it was shown that conventional ZF performs very poorly and several schemes were proposed to improve the robustness of

the transmission with respect to CSIT inconsistencies [4], [7], [8]. One of the main successes has been obtained for the 2-user setting where the DoF was shown to be equal to the DoF of the centralized setting [7] by means of an asymmetric precoding where some TX deliberately throws away instantaneous CSIT.

Yet, these works suffer from the limitations of the metric used, as DoF only provides the asymptotic rate *slope* with respect to the SNR. Since it does not provide any information about the beamforming gain or the efficient power use at the TXs, schemes resulting in the same DoF may need a considerably different power to achieve the same rate [9]. Hence, the natural next step towards capacity characterization is to study the *rate offset*, which is the constant term in the linear approximation of the sum rate at high SNR, i.e.,

$$R(P) = \text{DoF} \log_2(\text{SNR}) - \mathcal{L}_\infty + o(1) \tag{1}$$

where $\mathcal{L}_\infty$ represents the rate offset (vertical offset). Our main contributions read as follows:

- We provide a novel precoding scheme that achieves accurate ZF of the interference and, at the same time, high beamforming gain through consistent transmission at the TXs.
- Through a new lower bound, we show that the proposed scheme achieves a vanishing rate loss at high SNR when compared to the centralized setting with perfect CSI sharing.

Notations: We use the Landau notation, i.e., $f(x) = o(x)$ implies than $\lim_{x\to\infty} \frac{f(x)}{x} = 0$. $\mathbb{R}^+$ stands for $\{x \in \mathbb{R} : x > 0\}$, $\mathbb{E}_{|A}$ denotes the conditional expectation given an event A, and $\Pr(A)$ denotes the probability of an event A.

II. PROBLEM FORMULATION

A. Transmission Model

We consider a setting with 2 single-antenna TXs jointly serving 2 single-antenna RXs over a Network MIMO setting –also known as Distributed Broadcast Channel (BC)–. The extension to multiple-antenna TXs but single-antenna RXs is more challenging and is relegated to the journal version of this work. The signal received at RX i is

$$y_i = \mathbf{h}_i^{\text{H}} \mathbf{x} + z_i, \tag{2}$$

where $\mathbf{h}_i^{\text{H}} \in \mathbb{C}^{1\times 2}$ is the channel coefficients vector towards RX i, $\mathbf{x} \in \mathbb{C}^{2\times 1}$ is the transmitted multi-user signal, and $z_i \in \mathbb{C}$ is the Additive White Gaussian Noise (AWGN) at RX i, being independent of the channel and the transmitted

signal, and drawn from a circularly symmetric complex Gaussian distribution ($\mathcal{N}_{\mathbb{C}}(0, 1)$). We further define the channel matrix $\mathbf{H} \in \mathbb{C}^{2 \times 2}$ as

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{h}_1^{\mathrm{H}} \\ \mathbf{h}_2^{\mathrm{H}} \end{bmatrix}, \tag{3}$$

with its (i, k)-th element representing the channel coefficient from TX k to RX i and being denoted as h_{ik}. The channel coefficients are assumed to be i.i.d. as $\mathcal{N}_{\mathbb{C}}(0, 1)$ such that all the channel sub-matrices are full rank with probability one.

The transmitted multi-user signal $\mathbf{x} \in \mathbb{C}^{2 \times 1}$, is obtained from the precoding of the symbol vector $\boldsymbol{s} \triangleq [s_1 \ s_2]^{\mathrm{T}}$. The symbols s_i are i.i.d. as $\mathcal{N}_{\mathbb{C}}(0, 1)$ and s_i denotes the symbol intended by RX i such that

$$\mathbf{x} \triangleq \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}, \tag{4}$$

where P is the transmit power per TX. The vector $\mathbf{t}_i \in \mathbb{C}^{2 \times 1}$ denotes the precoding vector towards RX i. For further reference, we define $\mathbf{T} \triangleq \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 \end{bmatrix}$ as the multi-user precoder, and the precoder of TX j as $\mathbf{t}_{\mathrm{TX}\,j} \triangleq \begin{bmatrix} \{\mathbf{t}_1\}_j & \{\mathbf{t}_1\}_2 \end{bmatrix}^{\mathrm{T}}$. We assume a per-TX instantaneous power constraint *for the precoder*, i.e., $\|\mathbf{t}_{\mathrm{TX}\,j}\| \leq 1$, $\forall j \in \{1, 2\}$, such that $\mathbb{E}\big[\|\mathbf{x}\|^2\big] \leq P$.

B. Distributed CSIT Model

We consider in this work that the RXs have perfect channel knowledge to focus on the challenges of CSI feedback and limited CSI sharing among TXs. As previously mentioned, we consider here a Distributed CSIT configuration in which each TX receives a different imperfect estimate of the multi-user channel [4]. For sake of exposition, we consider that the CSI accuracy available at TX j is *homogeneous* across RXs. Note that our results are not restricted by this assumption and they extend to the case with different accuracy for each RX.

It is known that, in order to avoid the collapse of DoF in the Centralized CSIT setting, the CSIT error variance has to scale as $P^{-\alpha}$, with $\alpha > 0$, [1], [2], where α is called the *CSIT scaling coefficient*. Based on that result, we extend the model to the distributed setting by assuming that the error variance at TX j scales as $P^{-\alpha^{(j)}}$, with $\alpha^{(j)} > 0$ and $\alpha^{(1)} \neq \alpha^{(2)}$.

Specifically, we consider that RX i feeds back to TX j a quantized version of his normalized vector $\tilde{\mathbf{h}}_i \triangleq \mathbf{h}_i / \|\mathbf{h}_i\| \in \mathbb{C}^2$ using $B^{(j)}$ bits, denoted as $\hat{\mathbf{h}}_i^{(j)}$. After receiving the feedback from both RXs, TX j obtains a multi-user channel estimate $\hat{\mathbf{H}}^{(j)} = [\hat{\mathbf{h}}_1^{(j)}, \hat{\mathbf{h}}_2^{(j)}]^{\mathrm{H}} \in \mathbb{C}^{2 \times 2}$. Similar to [2], we assume that RX i uses random vector quantization codebooks of $2^{B^{(j)}}$ codewords and that the number of quantization bits grows linearly with $\log_2(P)$ as

$$B^{(j)} = \alpha^{(j)} \log_2(P). \tag{5}$$

This implies that the CSIT error variance at TX j scales as $P^{-\alpha^{(j)}}$ (since $P^{-\alpha^{(j)}} = 2^{-B^{(j)}}$ [2]). In order to avoid degenerate conditions, we assume that the codebooks of different RXs do not share any codeword. Moreover, we can order the TXs w.l.o.g. such that

$$1 \geq \alpha^{(1)} \geq \alpha^{(2)} > 0. \tag{6}$$

C. Figure-of-Merit

Our figure-of-merit is the expected sum rate over both the fading realizations and the random codebooks. Let us define the expected rate of RX i as $R_i \triangleq \mathbb{E}[r_i]$, where r_i is the instantaneous rate of RX i. In our setting, r_i writes as

$$r_i \triangleq \log_2 \left(1 + \frac{\frac{P}{2}|\mathbf{h}_i^{\mathrm{H}} \mathbf{t}_i|^2}{1 + \frac{P}{2}|\mathbf{h}_i^{\mathrm{H}} \mathbf{t}_{\bar{i}}|^2} \right), \tag{7}$$

where we have introduced the notation $\bar{i} \triangleq i \pmod 2 + 1$. Then, the expected sum rate is given by $R \triangleq R_1 + R_2$.

D. Centralized ZF Precoding

We restrict this work to ZF precoding schemes, which are known to achieve the optimal DoF in the centralized CSIT setting [1], [2] and that allow for analytical tractability. In this "ideal" centralized setting, all the TXs have access to the same channel estimate $\hat{\mathbf{H}}$. We consequently define $\hat{\mathbf{h}}_i$, α, as the centralized counterparts of $\hat{\mathbf{h}}_i^{(j)}$, $\alpha^{(j)}$, respectively. Let $\mathbf{v}_i^\star$ denote a unit-norm ZF precoder for RX i, computed on the basis of the estimate $\hat{\mathbf{H}}$. We can then write the centralized ZF precoding matrix as $\mathbf{T}^{\mathrm{ZF}} \triangleq \begin{bmatrix} \mu_1 \mathbf{v}_1^\star & \mu_2 \mathbf{v}_2^\star \end{bmatrix}$, where $\mu_i \in \mathbb{R}$ is a parameter that ensures that the instantaneous power constraint $\|\mathbf{t}_{\mathrm{TX}\,j}\| \leq 1$ is fulfilled. From the ZF precoding definition, $\mathbf{v}_i^\star$ is a vector satisfying that

$$\hat{\mathbf{h}}_{\bar{i}}^{\mathrm{H}} \mathbf{v}_i^\star = 0. \tag{8}$$

Given that multiplying the beamformer $\mathbf{v}_i^\star$ by a phase-shift $e^{\imath \phi_i}$ does not impact the rate [10], we can select w.l.o.g., among all the possible $\mathbf{v}_i^\star$, the vector $\mathbf{v}_i = e^{-\imath \phi_i^v}[\hat{\mathrm{h}}_{\bar{i}2}, -\hat{\mathrm{h}}_{\bar{i}1}]^{\mathrm{T}}$, where ϕ_i^v is the phase of the second coefficient ($\hat{\mathrm{h}}_{\bar{i}1}$). Thus,

$$\mathbf{T}^{\mathrm{ZF}} = \underbrace{\begin{bmatrix} \hat{\mathrm{h}}_{21}^{-1}\hat{\mathrm{h}}_{22} & \hat{\mathrm{h}}_{11}^{-1}\hat{\mathrm{h}}_{12} \\ -1 & -1 \end{bmatrix}}_{\triangleq \mathbf{V}^\star} \underbrace{\begin{bmatrix} \lambda_1^\star & 0 \\ 0 & \lambda_2^\star \end{bmatrix}}_{\triangleq \boldsymbol{\Lambda}^\star}, \tag{9}$$

where we have introduced the notation $\lambda_i^\star \triangleq \mu_i |\mathbf{v}_{i,2}|$. From the unitary power constraint it holds that $0 \leq \lambda_i^\star \leq 1$. Expression in (9) is just a rewriting of the conventional ZF matrix used in the literature [2], introduced to make the analogy with the distributed approach more explicit, such that we detach the interference-nulling part ($\mathbf{V}^\star$) and the power control ($\boldsymbol{\Lambda}^\star$).

Regarding the power normalization coefficient μ_i, it is obtained from an arbitrary algorithm satisfying the per-TX precoder power constraint $\|\mathbf{t}_{\mathrm{TX}\,j}\| \leq 1$, $\forall j \in \{1, 2\}$, and such that the probability density function of λ_i, denoted by f_{Λ_i}, is bounded away from infinity such that

$$\max_x f_{\Lambda_i}(x) \leq f_{\Lambda_i}^{\max} < \infty. \tag{10}$$

A more detailed discussion of μ_i can be found in the extended version [11].

III. MAIN RESULTS

Although ZF precoding schemes as the one described in Section II-D perform properly with centralized CSIT, their performance shrinks considerably on the distributed CSIT setting. This comes from the fact that the zero-forcing accuracy

is proportional to the worst quality among the TXs ($\alpha^{(2)}$ in our setting). Thus, conventional ZF does not achieve the centralized DoF. Furthermore, if TX 1 tries to estimate TX 2's CSI based on its own estimate it will incur in an estimation error proportional to $\alpha^{(2)}$. The solution proposed in DoF-achieving schemes [4], [7], [8] –i.e., that TX 2 precodes with a vector independent of its instantaneous CSI– also succumbs to the assumption of instantaneous power constraint for the precoding vector ($\|\mathbf{t}_{\mathrm{TX}\,j}\| \leq 1$), since a less practical average power constraint was considered. The only scheme achieving the optimal DoF is obtained from [4] where the transmit power scales in $P/\log(P)$. This leads to a very inefficient power normalization, and hence to a very poor rate offset ($\mathcal{L}_\infty$).

We present a distributed precoding scheme, coined *Hybrid Active-Passive ZF Precoding* (HAP), that precludes TX 2 from harming the performance. The key for attaining such result is an asymmetric ZF scheme and the quantization of the power control, that allows the TXs to be *consistent*.

A. Proposed Hybrid Active-Passive ZF Precoding

Let $\mathcal{Q}(\cdot)$ represent the output of an arbitrary quantizer $\mathcal{Q}$ satisfying that $\mathcal{Q}(x) \leq x$. The HAP precoder, denoted by $\mathbf{T}^{\mathrm{HAP}} \in \mathbb{C}^{2\times 2}$, is given by

$$\mathbf{T}^{\mathrm{HAP}} \triangleq \begin{bmatrix} (\hat{\mathrm{h}}_{21}^{(1)})^{-1}\hat{\mathrm{h}}_{22}^{(1)} & (\hat{\mathrm{h}}_{11}^{(1)})^{-1}\hat{\mathrm{h}}_{12}^{(1)} \\ -1 & -1 \end{bmatrix} \odot \begin{bmatrix} \mathcal{Q}(\lambda_1^{(1)}) & \mathcal{Q}(\lambda_2^{(1)}) \\ \mathcal{Q}(\lambda_1^{(2)}) & \mathcal{Q}(\lambda_2^{(2)}) \end{bmatrix} \quad (11)$$

where $\odot$ denotes the Hadamard (element-wise) product and $\lambda_i^{(j)}$ is the distributed counterpart of $\lambda_i^\star$ estimated at TX j. We observe that the first matrix is equal to the interference-nulling matrix $\mathbf{V}^\star$ in (9) based on the imperfect channel estimate $\hat{\mathbf{H}}^{(1)}$, and hence it is independent of the CSI of TX 2. Conversely, the second matrix needs to be computed at both TXs, and thus it differs from the centralized power normalization matrix $\mathbf{\Lambda}^\star$. The idea behind this separation is that the interference-nulling has to be extremely accurate, but it can be performed by a single TX, whereas the power normalization has to be done by both TXs, but it can be computed with a reduced precision, allowing the TXs to be consistent.

Since $\lambda_i^{(j)} \in [0, 1]$, we have that $\mathcal{Q}(\lambda_i^{(j)}) \in [0, 1]$. Moreover, we assume that it exists $M_{\mathcal{Q}} < \infty$ such that

$$\left| \mathbb{E}_{|\mathcal{Q}(x)>0}\left[\log_2\left(\mathcal{Q}(x)\right)\right] \right| \leq M_{\mathcal{Q}}, \tag{P0}$$

which is a technical assumption that is satisfied by any non-degenerate quantizer. The role of $\mathcal{Q}$ is to trade-off the accuracy of the power control with the consistency of the decision at the TXs, as the ZF orthogonality of (8) is preserved only if both TXs obtain the same quantized value –if $\mathcal{Q}(\lambda_i^{(1)}) = \mathcal{Q}(\lambda_i^{(2)})$–. In order to emphasize the relevance of the quantizer, we define Ω as the set of estimates $(\hat{\mathbf{H}}^{(1)}, \hat{\mathbf{H}}^{(2)})$ that ensure that the ZF orthogonality is not violated, excluding degenerate cases, i.e.,

$$\Omega \triangleq \left\{(\hat{\mathbf{H}}^{(1)}, \hat{\mathbf{H}}^{(2)}) \,\middle|\, \forall i \in \{1, 2\}, \, \mathcal{Q}(\lambda_i^{(1)}) = \mathcal{Q}(\lambda_i^{(2)}) \in \mathbb{R}^+\right\}. \tag{12}$$

In simple words, Ω encloses the cases when the TXs agree on the power normalization coefficients for both RXs and they are strictly positive. We further denote the complementary event

of Ω as Ω^{c} (the *inconsistent* cases). We proceed by introducing two important properties for the quantizers.

Definition 1 (Asymptotically Accurate Quantizers): A quantizer $\mathcal{Q}$ is said to be *asymptotically accurate* if

$$\lim_{P\to\infty} \mathcal{Q}(\lambda_i^{(j)}) = \lambda_i^{(1)} \quad \text{a.s.} \quad \forall i, j \in \{1, 2\}, \tag{P1}$$

where a.s. stands for *almost surely*.

Definition 2 (Asymptotically Consistent Quantizers): A quantizer $\mathcal{Q}$ is said to be *asymptotically consistent* if

$$\Pr\left(\Omega^{\mathsf{c}}\right) = o\left(\frac{1}{\log_2(P)}\right). \tag{P2}$$

Property (P2) implies that *inconsistent precoding* events are negligible in terms of asymptotic rate. In the following lemma, we exhibit one particular quantizer satisfying (P1)-(P2). Optimizing further this quantizer is crucial to good performance at finite SNR and its optimization is an ongoing research topic.

Lemma 1: Let $\mathcal{Q}_u$ be a uniform quantizer in the interval $[0, 1]$ with a step size of $P^{-\alpha^{(2)}/4}$, such that

$$\mathcal{Q}_u(x) \triangleq P^{\frac{-\alpha^{(2)}}{4}}\left\lfloor P^{\frac{\alpha^{(2)}}{4}}x \right\rfloor. \tag{13}$$

Then, $\mathcal{Q}_u$ satisfies properties (P0), (P1) and (P2).

Proof: The proof is lengthy and it is hence relegated to the extended version [11] due to space constraints. ∎

B. Main Results

Let us denote by $R^{\mathrm{HAP}}(\alpha^{(1)}, \alpha^{(2)})$ the expected sum rate achieved using HAP precoding in the Distributed CSIT setting with CSIT scaling quality $(\alpha^{(1)}, \alpha^{(2)})$. Similarly, we denote as $R^{\mathrm{ZF}}(\alpha^{(1)})$ the expected sum rate attained by the centralized ZF precoder of Section II-D on the basis of the estimate $\hat{\mathbf{H}}^{(1)}$, i.e., $R^{\mathrm{ZF}}(\alpha^{(1)})$ is the rate achieved on an ideal centralized setting where TX 2 is provided with a more accurate estimate. Accordingly, the rate gap between those settings is defined as

$$\Delta\mathrm{R} \triangleq R^{\mathrm{ZF}}(\alpha^{(1)}) - R^{\mathrm{HAP}}(\alpha^{(1)}, \alpha^{(2)}). \tag{14}$$

We can now state our main results.

Theorem 1: The rate gap of ZF precoding with Distributed CSIT is upper bounded by

$$\Delta\mathrm{R} \leq 2\mathbb{E}_{|\Omega}[\log_2(\Gamma_1)] + \Pr(\Omega^{\mathsf{c}}) R_{|\Omega^{\mathsf{c}}}^{\mathrm{ZF}}(\alpha^{(1)}), \tag{15}$$

where Ω is defined in (12), $\Gamma_1 \triangleq \left|\lambda_1^{(1)}/\mathcal{Q}(\lambda_1^{(1)})\right|^2$, and it holds that $R_{|\Omega^{\mathsf{c}}}^{\mathrm{ZF}}(\alpha^{(1)}) \leq 2\log_2(1 + P)$.

The proof is detailed in Section IV. This bound depends on the set Ω and thus on the quantizer selected. Intuitively, a "good" quantizer has to ensure a probability of *inconsistency* – $\Pr(\Omega^{\mathsf{c}})$– small. This can be done by enlarging the quantization step, what will make the first term bigger, as $\mathcal{Q}(\lambda_1^{(1)})$ needs to be as close to $\lambda_1^{(1)}$ as possible. This shows why finding the optimal quantizer is a challenging research topic. Nevertheless, there exist quantizers that behave asymptotically optimal, as stated in the following theorem.

Theorem 2: Let $\mathcal{Q}$ be an arbitrary quantizer satisfying (P0), (P1) and (P2). Then, taking the limit in Theorem 1 yields

$$\lim_{P\to\infty} \Delta\mathrm{R} \leq 0. \tag{16}$$

Proof: The proof follows from Theorem 1. First, note that the sum rate $R^{\text{ZF}}_{|\Omega^c}(\alpha^{(1)})$ is trivially bounded by twice the interference-free single-user rate to obtain

$$R^{\text{ZF}}_{|\Omega^c}(\alpha^{(1)}) \leq 2\log_2(1+P), \tag{17}$$

what together with property (P2) implies that

$$\Pr(\Omega^c) R^{\text{ZF}}_{|\Omega^c}(\alpha^{(1)}) = o(1). \tag{18}$$

Thus, it only remains to show that $\lim_{P\to\infty} \mathbb{E}_{|\Omega}[\log_2(\Gamma_1)] = 0$. From the definition of Γ_1 in Theorem 1, it holds that

$$\mathbb{E}_{|\Omega}[\log_2(\Gamma_1)] = \mathbb{E}_{|\Omega}\left[\log_2\left(\lambda_1^{(1)}\right)\right] - \mathbb{E}_{|\Omega}\left[\log_2\left(\mathcal{Q}(\lambda_1^{(1)})\right)\right]. \tag{19}$$

Note that, for any variable x such that $0 \leq x \leq 1$, and for any two events A, B, such that $0 < \Pr(B \mid A) < 1$, it holds that

$$\mathbb{E}_{|A}[\log_2(x)] = \Pr(B \mid A)\mathbb{E}_{|A\cap B}[\log_2(x)]$$
$$+ \Pr(B^c \mid A)\mathbb{E}_{|A\cap B^c}[\log_2(x)]. \tag{20}$$

Since $0 \leq x \leq 1$, $\mathbb{E}_{|A\cap B^c}[\log_2(x)] \leq 0$ and hence

$$\mathbb{E}_{|A\cap B}[\log_2(x)] \geq \frac{1}{\Pr(B \mid A)}\mathbb{E}_{|A}[\log_2(x)]. \tag{21}$$

Therefore, if $\mathbb{E}_{|A}[\log_2(x)]$ exists, also $\mathbb{E}_{|A\cap B}[\log_2(x)]$ exists and it is bounded below by (21) and above by 0. Let A and B be $A = \{\mathcal{Q}(\lambda_i^{(1)}) > 0, \forall i\}$ and $B = \{\mathcal{Q}(\lambda_i^{(1)}) = \mathcal{Q}(\lambda_i^{(2)}), \forall i\}$. Thus, $\Omega = A \cap B$. It follows from (21) and (P0) that

$$\mathbb{E}_{|\Omega}[\log_2(\mathcal{Q}(\lambda_1^{(1)}))] \geq -\frac{\Pr(\mathcal{Q}(\lambda_i^{(1)}) > 0, \forall i)}{\Pr(\Omega)}M_{\mathcal{Q}}, \tag{22}$$

where we have applied the fact that $\Pr(B|A) = \frac{\Pr(A\cap B)}{\Pr(A)}$. Hence, $\mathbb{E}_{|\Omega}[\log_2(\mathcal{Q}(\lambda_1^{(1)}))]$ is bounded. The same result follows for $\mathbb{E}_{|\Omega}[\log_2(\lambda_1^{(1)})]$ from the bounded density assumption of (10). Moreover, from the continuity of the $\log$ function and (P1), $\log_2(\mathcal{Q}(\lambda_1^{(1)}))$ converges a.s. to $\log_2(\lambda_1^{(1)})$. From all these facts, we can apply Lebesgue's Dominated Convergence Theorem [12, Theorem 16.4] to show that

$$\lim_{P\to\infty} \mathbb{E}_{|\Omega}\left[\log_2\left(\mathcal{Q}(\lambda_1^{(1)})\right)\right] = \mathbb{E}_{|\Omega}\left[\log_2\left(\lambda_1^{(1)}\right)\right], \tag{23}$$

and thus $\lim_{P\to\infty} \mathbb{E}_{|\Omega}[\log_2(\Gamma_1)] = 0$, which concludes the proof. ∎

Corollary 1 (Rate Offset with HAP precoder): It holds from Theorem 2 that the rate offset $\mathcal{L}_\infty$ –defined in (1)– of ZF with distributed CSIT is the same as for the genie-aided centralized setting, whose rate offset was shown in [2] to be constant with respect to Perfect CSIT ZF (and thus with respect to the capacity-achieving Dirty Paper Coding) for $\alpha = 1$.

The key for attaining such surprising performance is the trade-off between *consistency* and *accuracy* that is ruled by the quantizer. Interestingly, Lemma 1 illustrates that simple quantizers –as the uniform one– satisfy the sufficient conditions of convergence if we select the correct number of quantization levels. Moreover, since this quantizer is applied locally and no information exchange is done, the granularity of the quantizer does not increase the complexity of the scheme.

Let us consider that there is agreement between the TXs, i.e., that $\mathcal{Q}(\lambda_i^{(1)}) = \mathcal{Q}(\lambda_i^{(2)}), \forall i \in \{1, 2\}$, such that we define

$$\lambda_i^{\mathcal{Q}} \triangleq \mathcal{Q}(\lambda_i^{(j)}), \qquad \forall i, j \in \{1, 2\}, \tag{24}$$

what (P2) ensures that occurs with a probability high enough such that the disagreement is asymptotically negligible. In this case we can rewrite (11) as

$$\mathbf{T}^{\text{HAP}} \triangleq \begin{bmatrix} (\hat{h}_{21}^{(1)})^{-1}\hat{h}_{22}^{(1)} & (\hat{h}_{11}^{(1)})^{-1}\hat{h}_{12}^{(1)} \\ -1 & -1 \end{bmatrix} \begin{bmatrix} \lambda_1^{\mathcal{Q}} & 0 \\ 0 & \lambda_2^{\mathcal{Q}} \end{bmatrix}. \tag{25}$$

It becomes then clear that the orthogonality (i.e., the interference attenuation) is ensured by the first matrix in (25) while the second diagonal matrix is only used to satisfy the power constraint. Regarding the quantizer $\mathcal{Q}$, note that letting $\mathcal{Q}$ have a single quantization point leads to a statistical power control, whereas letting $\mathcal{Q}$ have infinite points leads to the unquantized version. In both cases, part of the DoF is lost.

IV. PROOF OF THEOREM 1

We consider w.l.o.g. the rate difference at RX 1, denoted as ΔR_1, since the proof for RX 2 is obtained after switching the indices of the RXs. ΔR_1 can be split as

$$\Delta R_1 = \Pr(\Omega)\Delta R_{1|\Omega} + \Pr(\Omega^c)\Delta R_{1|\Omega^c}. \tag{26}$$

First, we focus on $\Delta R_{1|\Omega}$, which encloses the *consistent precoding* cases. Conditioned on Ω it holds that $\mathcal{Q}(\lambda_i^{(1)}) = \mathcal{Q}(\lambda_i^{(2)})$, $\forall i \in \{1, 2\}$, and hence we can use the notation $\lambda_i^{\mathcal{Q}}$ introduced in (24). Moreover, it can be observed from (9) and (25) that, conditioned on Ω, the HAP precoder satisfies

$$\mathbf{t}_i^{\text{HAP}} = \frac{\lambda_i^{\mathcal{Q}}}{\lambda_i^\star}\mathbf{t}_i^{\text{ZF}}, \qquad \forall i \in \{1, 2\}. \tag{27}$$

Since we assume that in the centralized ZF setting both TXs share the channel estimate of TX 1 ($\hat{\mathbf{H}}^{(1)}$), we have that $\lambda_i^\star = \lambda_i^{(1)}$. Thus, since $\mathcal{Q}(x) \leq x$, it follows that $\lambda_i^{\mathcal{Q}}/\lambda_i^\star \leq 1$, $\forall i \in \{1, 2\}$. Let us recall that Γ_i is defined as

$$\Gamma_i \triangleq \left|\frac{\lambda_i^{(1)}}{\lambda_i^{\mathcal{Q}}}\right|^2. \tag{28}$$

Conditioned on Ω, we can write that the SINR obtained through HAP precoding satisfies

$$1 + \frac{\frac{P}{2}|\mathbf{h}_1^{\text{H}}\mathbf{t}_1^{\text{HAP}}|^2}{1 + \frac{P}{2}|\mathbf{h}_1^{\text{H}}\mathbf{t}_2^{\text{HAP}}|^2} = 1 + \frac{\frac{1}{\Gamma_1}\frac{P}{2}|\mathbf{h}_1^{\text{H}}\mathbf{t}_1^{\text{ZF}}|^2}{1 + \frac{1}{\Gamma_2}\frac{P}{2}|\mathbf{h}_1^{\text{H}}\mathbf{t}_2^{\text{ZF}}|^2} \tag{29}$$

$$\geq \frac{1}{\Gamma_1}\left(1 + \frac{\frac{P}{2}|\mathbf{h}_1^{\text{H}}\mathbf{t}_1^{\text{ZF}}|^2}{1 + \frac{P}{2}|\mathbf{h}_1^{\text{H}}\mathbf{t}_2^{\text{ZF}}|^2}\right), \tag{30}$$

where (29) follows from (27)-(28) whereas (30) comes from the fact that $1/\Gamma_i \leq 1$, $\forall i$. We can recognize in (30) the SINR at RX 1 for the centralized ZF scheme such that it holds

$$R_{1|\Omega}^{\text{HAP}}(\alpha^{(1)}, \alpha^{(2)}) = \mathbb{E}_{|\Omega}\left[\log_2\left(1 + \frac{\frac{P}{2}|\mathbf{h}_1^{\text{H}}\mathbf{t}_1^{\text{HAP}}|^2}{1 + \frac{P}{2}|\mathbf{h}_1^{\text{H}}\mathbf{t}_2^{\text{HAP}}|^2}\right)\right] \tag{31}$$

$$\geq -\mathbb{E}_{|\Omega}[\log_2(\Gamma_1)] + R_{1|\Omega}^{\text{ZF}}(\alpha^{(1)}). \tag{32}$$

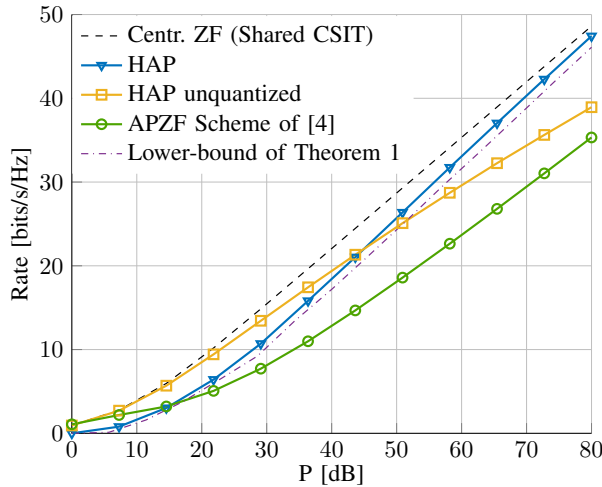Fig. 1: Expected sum rate for the setting with CSIT scaling parameters $\alpha^{(1)} = 1$, $\alpha^{(2)} = 0.6$, using the quantizer of Lemma 1.

Since $\Delta R_{1|\Omega} = R_{1|\Omega}^{\mathrm{ZF}}(\alpha^{(1)}) - R_{1|\Omega}^{\mathrm{HAP}}(\alpha^{(1)}, \alpha^{(2)})$, it follows that

$$\Delta R_{1|\Omega} \leq \mathbb{E}_{|\Omega}[\log_2(\Gamma_1)]. \tag{33}$$

Focusing on the *inconsistent precoding* cases, the rate gap can be bounded by the centralized rate as $\Delta R_{1|\Omega^c} \leq R_{1|\Omega^c}^{\mathrm{ZF}}(\alpha^{(1)})$. Putting these results together in (26) yields

$$\Delta R_1 \leq \mathbb{E}_{|\Omega}\left[2\log_2\left(\frac{\lambda_i^{(1)}}{\lambda_i^{\mathcal{Q}}}\right)\right] + \Pr(\Omega^c) R_{1|\Omega^c}^{\mathrm{ZF}}(\alpha^{(1)}), \tag{34}$$

where we have applied the fact that $\Pr(\Omega) \leq 1$. Thus, since Γ_1 and Γ_2 are identically distributed, it holds that $\Delta R \leq 2\Delta R_1$, which concludes the proof.

V. NUMERICAL RESULTS

We illustrate in the following the performance for the uniform quantizer $\mathcal{Q}_u$ introduced in Lemma 1. For sake of exposition, we assume a simple power normalization that ensures the per-TX power constraint. Let us introduce the precoding vector of TX j before normalization as $\mathbf{v}_{\mathrm{TX}\,j} = [\mathbf{v}_{1,j}, \mathbf{v}_{2,j}]^{\mathrm{T}}$, such that the final precoder of TX j is $\mathbf{t}_{\mathrm{TX}\,j} = [\mu_1 \mathbf{v}_{1,j}, \mu_2 \mathbf{v}_{2,j}]^{\mathrm{T}}$. Then, μ_i is chosen as

$$\mu_i \triangleq \frac{1}{\max(\|\mathbf{v}_{\mathrm{TX}\,1}\|, \|\mathbf{v}_{\mathrm{TX}\,2}\|)} \qquad \forall i \in \{1, 2\}. \tag{35}$$

In Fig. 1, we simulate the expected sum rate of the proposed scheme using Monte-Carlo runs and averaging over 1000 random codebooks and 1000 channel realizations, for the CSIT configuration $\alpha^{(1)} = 1$ and $\alpha^{(2)} = 0.6$. We can see that the proposed scheme leads to a vanishing loss with respect to the centralized case (where both TXs are provided with the best CSIT), and that the lower-bound of Theorem 1 is considerably close to the actual rate. Furthermore, the scheme given in [4] using a scaled power normalization of $P/\log_2(P)$

–so as to guarantee a full DoF and an instantaneous power constraint for the precoder $\mathbf{t}_{\mathrm{TX}\,j}$– can be seen to achieve also the optimal DoF although at the cost of a strong loss in rate offset. Finally, we can see how using an unquantized coefficient at TX 2 leads to a loss in terms of DoF. This occurs because, as aforementioned, the mismatches between the precoding coefficients of each TX break the orthogonality needed for the interference nulling. Thus, this scheme only achieves a DoF proportional to $\alpha^{(2)}$ instead of $\alpha^{(1)}$. At intermediate SNR, this unquantized scheme outperforms the proposed HAP scheme. Yet, it is a consequence of our focus in this work towards analytical tractability and asymptotic analysis. Optimizing the precoder for finite SNR performance will allow to bridge the gap between the two schemes to obtain a scheme outperforming both of them.

VI. CONCLUSION

Considering a decentralized scenario where each TX has a CSI with different SNR scaling accuracy, we have shown that there exists a linear precoding scheme that asymptotically recovers the rate of ZF precoding in the ideal centralized setting in which the best estimate is shared. Going beyond the setting considered, we have shown how using a low rate quantization of some parameters (here the power normalization) in combination with a higher-accuracy distributed decision allows to reach coordination without loosing precision. The extension of the results to more antennas and more users, as well as the optimization at finite SNR, are interesting and challenging research problems currently under investigation.

REFERENCES

[1] A. G. Davoodi and S. A. Jafar, "Aligned Image Sets Under Channel Uncertainty: Settling Conjectures on the Collapse of Degrees of Freedom Under Finite Precision CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5603–5618, Oct 2016.

[2] N. Jindal, "MIMO Broadcast Channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[3] D. Gesbert and P. de Kerret, "Team Methods for Device Cooperation in Wireless Networks," in *Cooperative and Graph Signal Processing*, ch. 18, pp. 469 – 487, Ed. by P. M. Djurić and C. Richard, Academic Press, 2018.

[4] P. de Kerret and D. Gesbert, "Degrees of freedom of the network MIMO channel with distributed CSI," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6806–6824, Nov. 2012.

[5] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of Caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[6] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

[7] A. Bazco-Nogueras, P. de Kerret, D. Gesbert, and N. Gresset, "Distributed CSIT Does Not Reduce the Generalized DoF of the 2-user MISO Broadcast Channel," *IEEE Wireless Commun. Lett.*, 2018.

[8] ——, "On the Degrees-of-Freedom of the K-user Distributed Broadcast Channel," 2018, submitted to IEEE Trans. Inf. Theory.

[9] A. Lozano, A. M. Tulino, and S. Verdu, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec 2005.

[10] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, 2008.

[11] A. Bazco-Nogueras, L. Miretti, P. de Kerret, D. Gesbert, and N. Gresset, "Achieving Vanishing Rate Loss in Decentralized Network MIMO," 2019. [Online]. Available: https://arxiv.org/abs/1901.06849

[12] P. Billingsley, *Probability and Measure*, ser. Wiley Series in Probability and Statistics. Wiley, 1995.